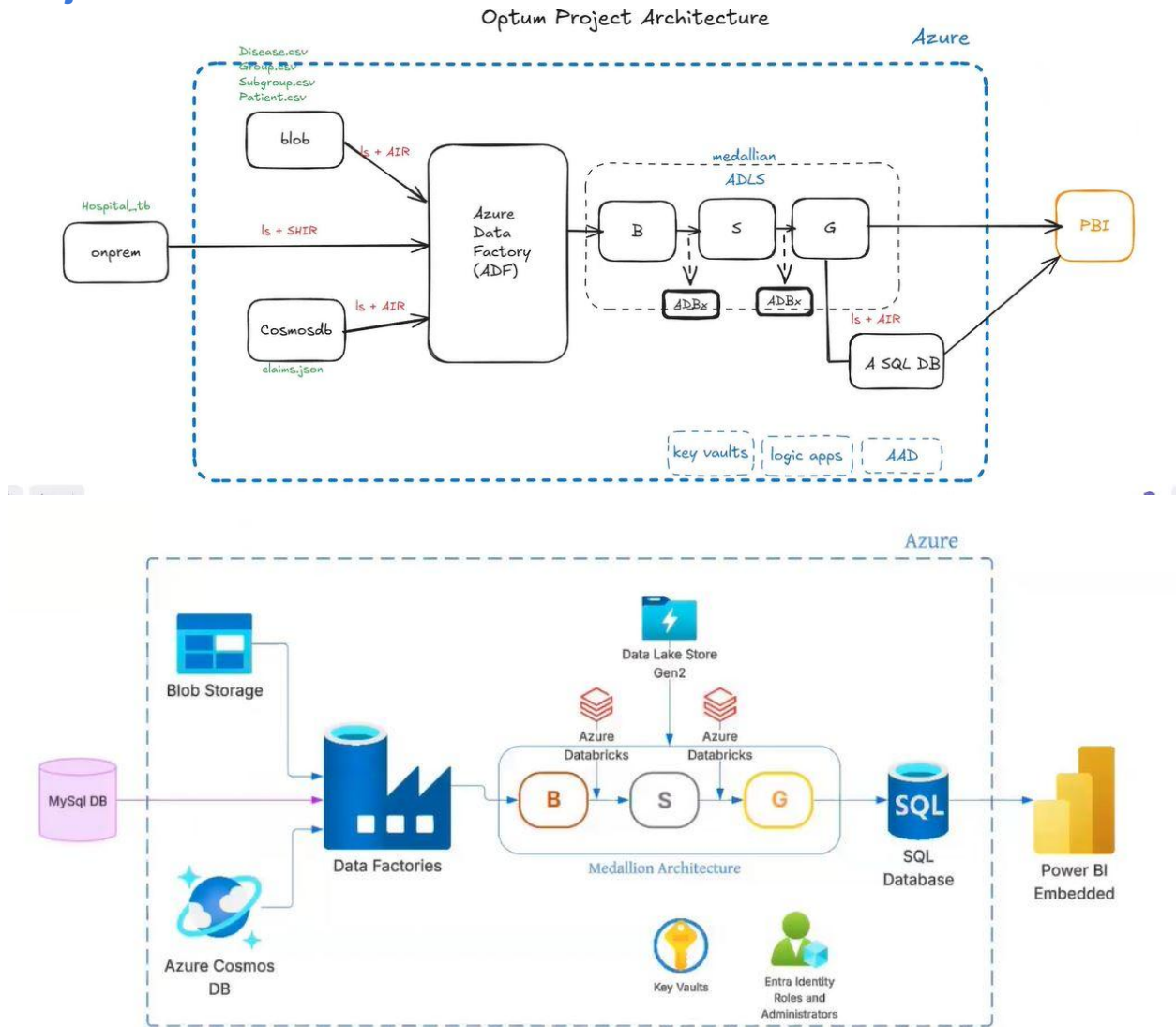


Optum Project

Project Objective:

This project aims to centralize data from multiple sources including MySQL, Azure Cosmos DB, and Blob Storage, enabling scalable and reliable data ingestion and transformation. It implements a Medallion Architecture (Bronze, Silver, and Gold) to ensure data quality, consistency, and governance.

Project Architecture:



Data Sources:

- **MySQL** : Hospital_TB
- **Azure CosmosDB** : claims.json
- **Azure Blob Storage** : disease.csv, group.csv, subgroup.csv, patient.csv, subscriber_record.csv

Data Ingestion and Orchestration:

- Azure Data Factory

Data Processing:

- Databricks

Data Serving Layer:

- Azure Database

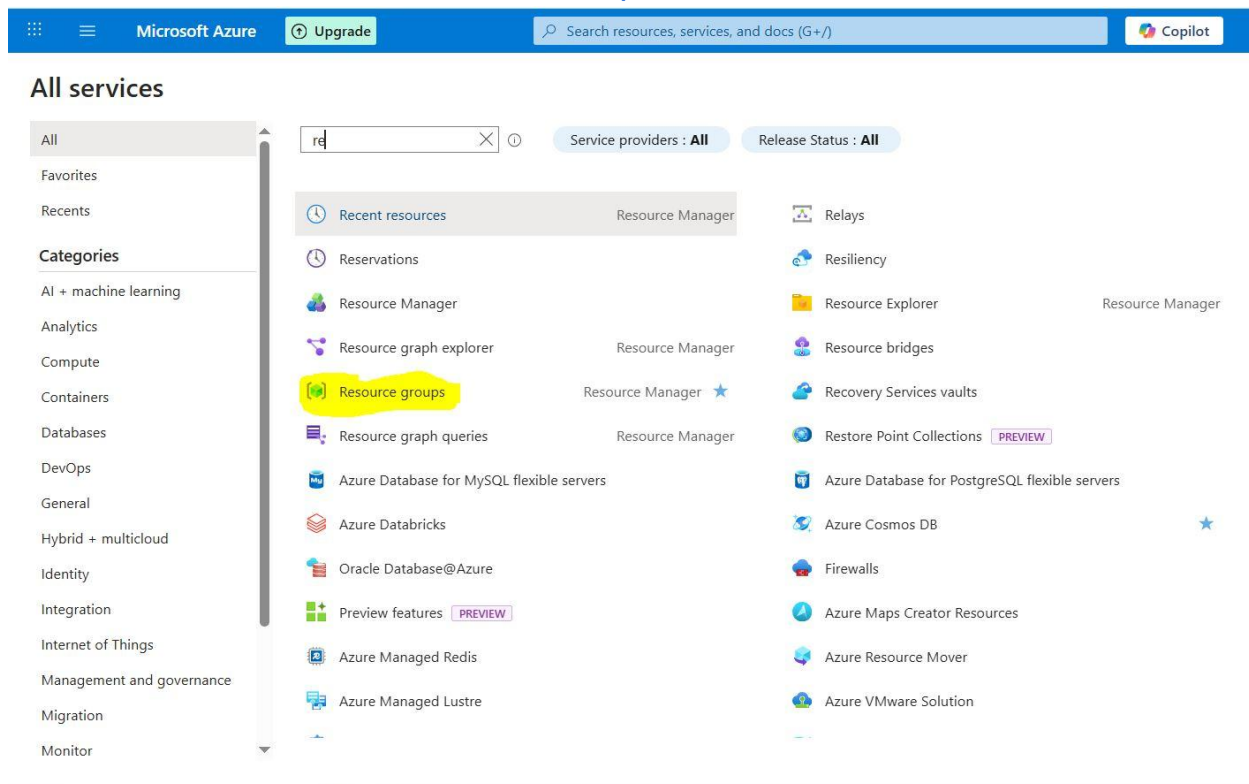
Creating services:

1.Resource Group

Resource group is like a Single Folder to keep all the related azure services for the particular project.

1.1.How to Create Resource Group

In Azure, Click Resource Group and Create it.



Create a resource group ...

Basics

Tags

Review + create

Resource group - A container that holds related resources for an Azure solution. The resource group can include all the resources for the solution, or only those resources that you want to manage as a group. You decide how you want to allocate resources to resource groups based on what makes the most sense for your organization. [Learn more](#)

Subscription * ⓘ

(Disabled) Azure subscription 1

Resource group name * ⓘ

Project_RG

Region * ⓘ

(US) East US

Previous

Next

Review + create

2.Azure Blob Storage

Create Azure blob and load the datasets

Create a resource



[Design a new Azure workload](#)

[Help me build a new Azure C](#)

Get Started

[Recently created](#)

Categories

[Machine Learning](#)

[AI Apps and Agents](#)

[Analytics](#)

[Blockchain](#)

[Infrastructure Services](#)

[Databases](#)

[Developer Tools](#)

[DevOps](#)

[Identity](#)

[Integration](#)

[Internet of Things](#)

[IT & Management Tools](#)

[Media](#)



G

Popular Azure services [See more in All services](#)



Function App

[Create](#)



Web App

[Create](#)



Virtual network

[Create](#)



Resource group

[Create](#)



Key Vault

[Create](#)



Virtual machine

[Create](#)



Storage account

[Create](#)

optumdatas

Container

Search

+ Add Directory

Upload

Change access level

Refresh

Delete

Copy

Paste

Rename

Acquire lease

Break lease

Edit columns

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

Shared access tokens

Access policy

Properties

Metadata

optumdatas

Authentication method: Access key (Switch to Microsoft Entra user account)

Add filter

Search blobs by prefix (case-sensitive)

Only show active blobs

Showing all 5 items

<input type="checkbox"/>	Name	Last modified	Access tier	Blob type	Size	Lease state
<input type="checkbox"/>	Patient_records.csv	12/6/2025, 10:19:02 PM	Hot (Inferred)	Block blob	4.99 KiB	Available
<input type="checkbox"/>	disease.csv	12/6/2025, 10:19:02 PM	Hot (Inferred)	Block blob	1.45 KiB	Available
<input type="checkbox"/>	group.csv	12/6/2025, 10:19:02 PM	Hot (Inferred)	Block blob	4.29 KiB	Available
<input type="checkbox"/>	subgroup.csv	12/6/2025, 10:19:02 PM	Hot (Inferred)	Block blob	561 B	Available
<input type="checkbox"/>	subscriber.csv	12/6/2025, 11:17:15 PM	Hot (Inferred)	Block blob	11.78 KiB	Available

3. Azure datalake storage gen2

1. Create ADs to create multiple directory for organising the data in a medallion architecture. ADs is in actual blob storage which needs to enable 'Hierarchical namespace' while creating it.

Create a storage account

Allow enabling anonymous access on individual containers

Enable storage account key access

Default to Microsoft Entra authorization in the Azure portal

Minimum TLS version

Permitted scope for copy operations (preview)

Hierarchical Namespace

Enable hierarchical namespace

Access protocols

Enable SFTP

Enable network file system v3

Version 1.2

From any storage account

Hierarchical namespace, complemented by Data Lake Storage Gen2 endpoint, enables file and directory semantics, accelerates big data analytics workloads, and enables access control lists (ACLs) [Learn more](#)

Enable hierarchical namespace

Blob and Data Lake Gen2 endpoints are provisioned by default [Learn more](#)

2. Create Directories as Medallion, in medallion folder, create 3 folders, 'Bronze', 'Silver', 'Gold'.

optumproadlsstorage

Storage account

Search

+ Add container

Upload

Refresh

Delete

Change access level

Restore containers

Edit columns

Overview

Activity log

Tags

Diagnose and solve problems

Access Control (IAM)

Data migration

Events

Storage browser

Partner solutions

Resource visualizer

Data storage

Search containers by prefix

Only show active containers

Showing all 1 items

All services > optumproadstorage_1/65141016/13 | Overview > optumproadstorage | Containers >

medallion ...

Container

Search

+ Add Directory | Upload | Refresh | Delete | Copy | Paste | Rename | Acquire lease | Break lease | Edit columns

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

medallion

Authentication method: Access key (Switch to Microsoft Entra user account)

Search blobs by prefix (case-sensitive)

Only show active objects

Showing all 3 items

<input type="checkbox"/>	Name	Last modified	Access tier	Blob type	Size	Lease state
<input type="checkbox"/>	Bronze	12/8/2025, 2:29:29 AM				...
<input type="checkbox"/>	Gold	12/8/2025, 2:29:37 AM				...
<input type="checkbox"/>	Silver	12/8/2025, 2:29:44 AM				...

3.Azure Data Factory

1.Create adf for orchestration and data ingestion.

2.Create from azure services

Microsoft Azure | Data Factory | optumproadstorage_1/65141016/13 | Search

Data Factory | Validate all | Publish all

Factory Resources

Filter resources by name

- Pipelines 0
- Change Data Capture (preview) 0
- Datasets 0
- Data flows 0
- Power Query 0

Select an item

Use the resource explorer to select or create a new item

All services

Data factories

Default Directory (ragavinadhi17outlook.onmicrosoft.com)

Identify non-compliant Data factories in my environment.

Sort Data factories by creation date.

Group Data factories by location in ARG query.

+ Create | Manage view | Refresh | Export to CSV | Open query | Assign tags | Add to service group

You are viewing a new version of Browse experience. Click here to access the old experience.

Filter for any field... | Subscription equals all | Type equals all | Resource Group equals all | Location equals all | Add filter



No data factories to display

Integrate data silos with Azure Data Factory, a service built for all data integration needs and skill levels. Easily construct ETL and ELT processes code-free within the intuitive visual environment.

+ Create

[Learn more](#)

4.Azure CosmosDB

- 1.Create Cosmosdb from azure services
- 2.Select Azure cosmos db for nosql
- 3.select request unit(RI) database account

[Home](#) >

Create a resource



I need a new low-cost VM

Help me build a new Azure OpenAI applicati

Get Started

Recently created

Categories

Machine Learning

AI Apps and Agents

Analytics

Blockchain

Infrastructure Services

Databases

Developer Tools

DevOps

Identity

Integration

Internet of Things

IT & Management Tools

Media



Getting starte

Popular Azure services [See more in All services](#)



Azure Cosmos DB

[Create](#)



SQL Database

[Create](#)



Azure Database for PostgreSQL Flexible Server

[Create](#)



Azure SQL

[Create](#)



Azure Synapse Analytics

[Create](#)



Azure Cache for Redis

[Create](#)



Azure Database for MySQL Flexible Server

[Create](#)

Create an Azure Cosmos DB account ...

Which API best suits your workload?

Azure Cosmos DB is a fully managed NoSQL and relational database service for building scalable, high performance applications. [Learn more](#)

To start, select the API to create a new account. The API selection cannot be changed after account creation.

Recommended APIs Others

Azure Cosmos DB for NoSQL

Azure Cosmos DB's core, or native API for working with documents. Supports fast, flexible development with familiar SQL query language and client libraries for .NET, JavaScript, Python, and Java.

Create

[Learn more](#)

Azure DocumentDB (with MongoDB compatibility)

Fully managed database service for apps written for MongoDB. Recommended if you have existing MongoDB workloads that you plan to migrate to Azure Cosmos DB.

Create

[Learn more](#)

Give Feedback

 [Help improve this page](#)

Choose Architecture ...

Which type of resource?

Azure DocumentDB (with MongoDB Compatibility) is designed for modern workloads, AI/Vector search, and cost efficiency. Cosmos DB for MongoDB supports requests. [users pick Azure DocumentDB.](#)

Note: Once a resource is created, your choice cannot be changed.

Recommended: Azure DocumentDB

- Vector search + GenAI support
- Cost-efficient scaling model
- Open-source engine [Learn more](#)
- Designed for multi-cloud and hybrid cloud solutions
- [See documentation and supported features](#)

Create

Request unit (RU) database account

- Designed for point-reads and simple queries
- Instant, granular autoscaling
- Multi-region writes
- Limited for analytics and AI workloads
- [See documentation and supported features](#)

Create

4. Open Local MongoDB and try connecting with CosmosDB with String Connections.

5. In data explorer, click Connect, inside it, copy 'Primary Connection String'

6. paste the connection string in mongodb so that, we can connect the datas from mongodb.

All services > Microsoft Azure Cosmos DB-20251206222420 | Overview > optumcosmosfornosql

optumcosmosfornosql | Data Explorer ☆ ...

Azure Cosmos DB for MongoDB account (RU)

Search

Enable Azure Synapse Link Visual Studio Code

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Quick start

Data Explorer

Migrate to Azure DocumentDB

Resource visualizer

Settings

Integrations

Collections

Monitoring

Automation

Help

+ New Collection

Home

Connect

Welcome to Azure Cosmos DB

Globally distributed, multi-model database service for any scale

Launch quick start

Launch a quick start tutorial to get started with sample data

New Collection

Create a new container for storage and throughput

Connect

Prefer using your own choice of tooling? Find the connection string you need to connect

Recents

Top 3 things you need to know

[What is the MongoDB API?](#)
Understand Azure Cosmos DB for MongoDB and its features.

Learning Resources

[Data Explorer keyboard shortcuts](#)
Learn keyboard shortcuts to navigate Data Explorer.

All services > Microsoft Azure Cosmos DB-20251206222420 | Overview > optumcosmosfornosql

optumcosmosfornosql | Data Explorer ☆ ...

Azure Cosmos DB for MongoDB account (RU)

Search

Enable Azure Synapse Link Visual Studio Code

Overview

Activity log

Access control (IAM)

Tags

Diagnose and solve problems

Quick start

Data Explorer

Migrate to Azure DocumentDB

Resource visualizer

Settings

Integrations

Collections

Monitoring

Automation

Help

+ New Collection

Home

Connect

URI

<https://optumcosmosfornosql.documents.azure.com:443/>

Read-write Keys **Read-only Keys**

PRIMARY KEY

36oyysT9c4NsqDWk1u7Uq66oVrGZcJreTNSHMmHVWnybRzIQvWWZV3iZa30zcP0tfxQJMN843yACDb6OC2Lg==

SECONDARY KEY

PRIMARY CONNECTION STRING

SECONDARY CONNECTION STRING

[Download sample app](#)

Optum

Manage your connection settings

While connected, you may only personalize your connection's name, color or favorite status. To fully configure it, you must first disconnect. Beware that disconnecting might cause work in progress to be lost.

[Disconnect](#)

URI

mongodb://localhost:27017/AccountEndpoint=https://optumcosmosfornosql.documents.azure.com:443/;AccountKey=36oyysT9c4NsqDWk1u7Uq66oVrGZcJreTNSHMmHVWnybRzIQvWWZV3iZa30zcP0tfxQJMN843yACDb6OC2Lg=;

Name

Optum

Color

No Color

☐ **Favorite this connection**

Favoriting a connection will pin it to the top of your list of connections

[Cancel](#) [Save](#)

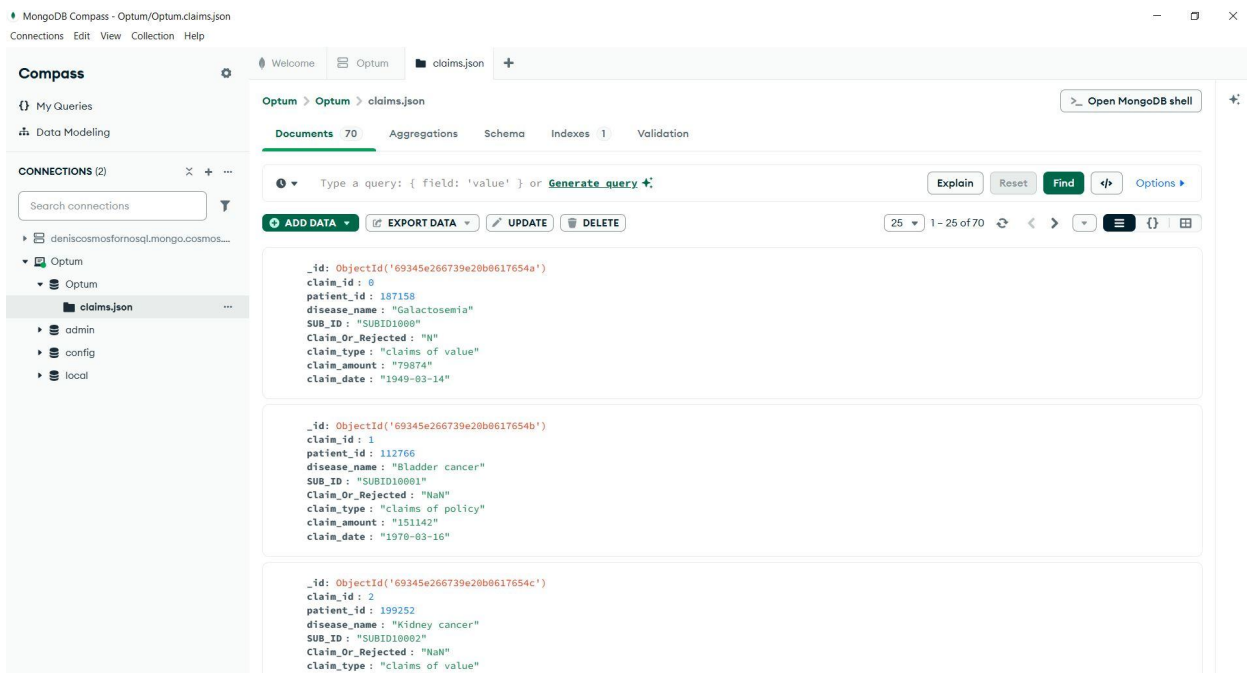
How do I find my connection string in Atlas?

If you have an Atlas cluster, go to the Cluster view. Click the 'Connect' button for the cluster to which you wish to connect.

[See example](#)

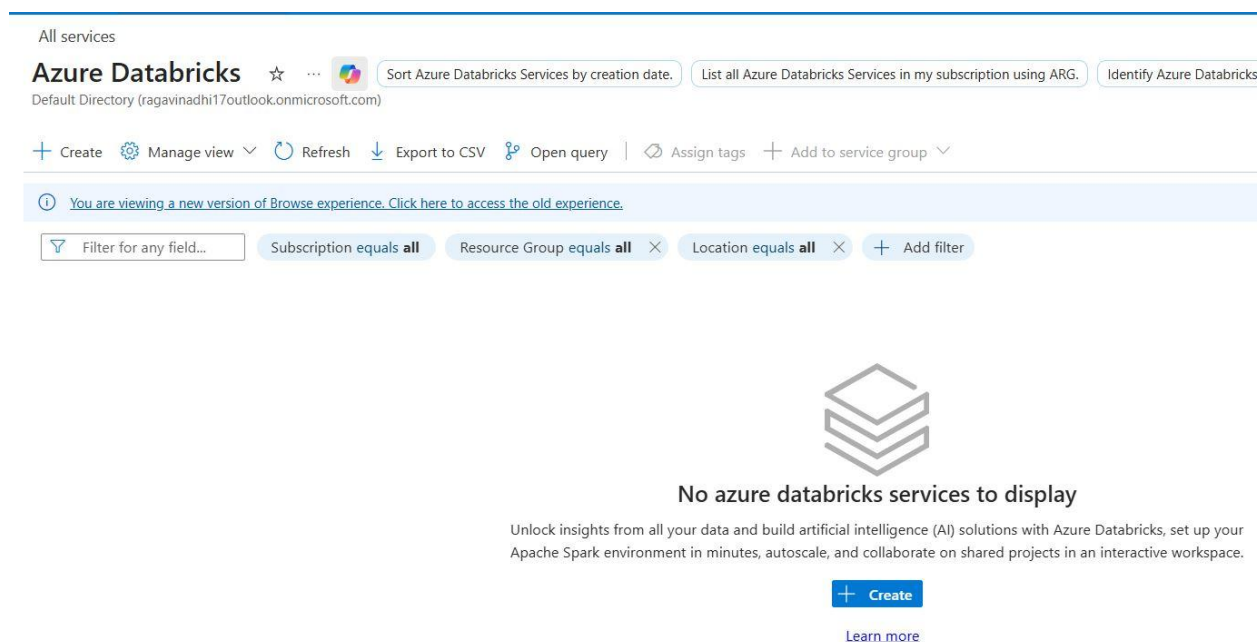
How do I format my connection string?

[See example](#)



5. Azure Databricks

1. Create Databricks from azure services to process/clean/transform the datas.



6. Azure SqlDB

1. Create SQL DB from azure services which is going to be centralized database to store the final denormalized table.

Fill in the areas;

- Servername : Optumsqlldb
- Location : East Asia
- Authentication method : use SQL authentication

- Password : *****
- Confirm Password : *****

2.Choose Standard or Basic Pricing Tier(for low cost)

- Backup Storage Redundancy : Locally Redundant backup Storage

3.Enable public Access(Firewall Rule):

- Connectivity method : Public Endpoint

4.Review and create

Creating link services:

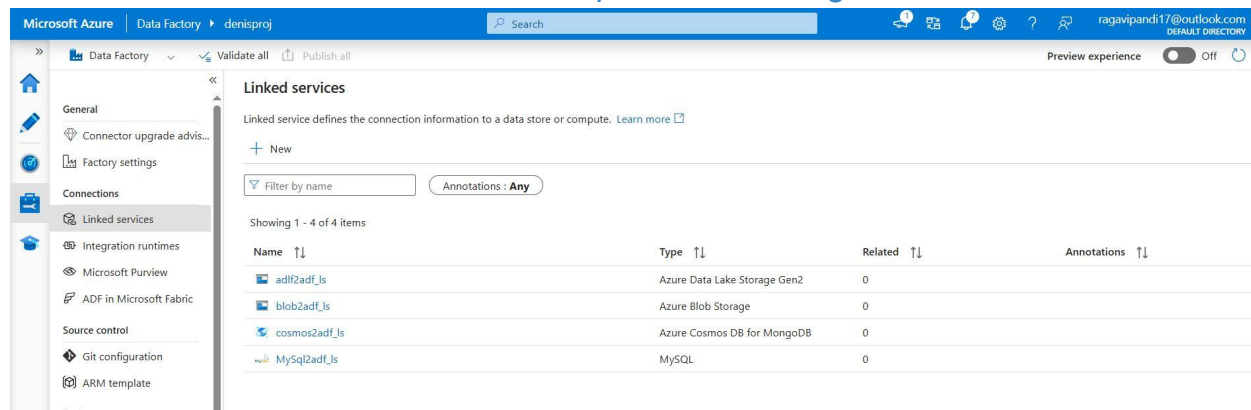
1.adls2adf_ls

2.blob2adf_ls

3.cosmosdb2adf_ls

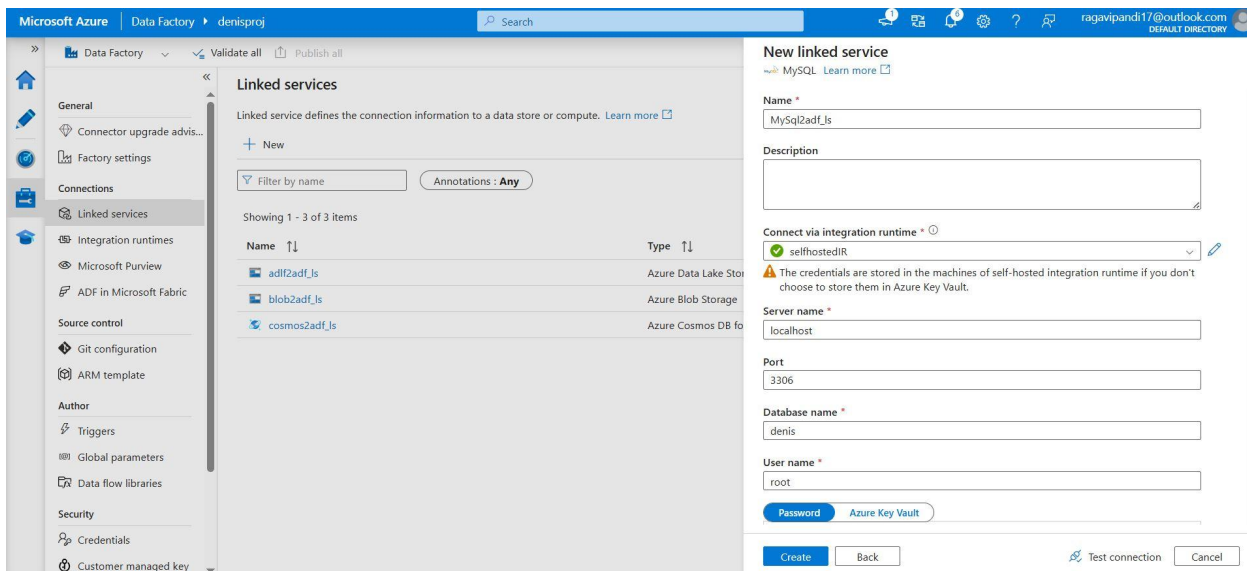
4.mysql2adf_ls

a.Create all linked services, it is like a path for data ingestion.

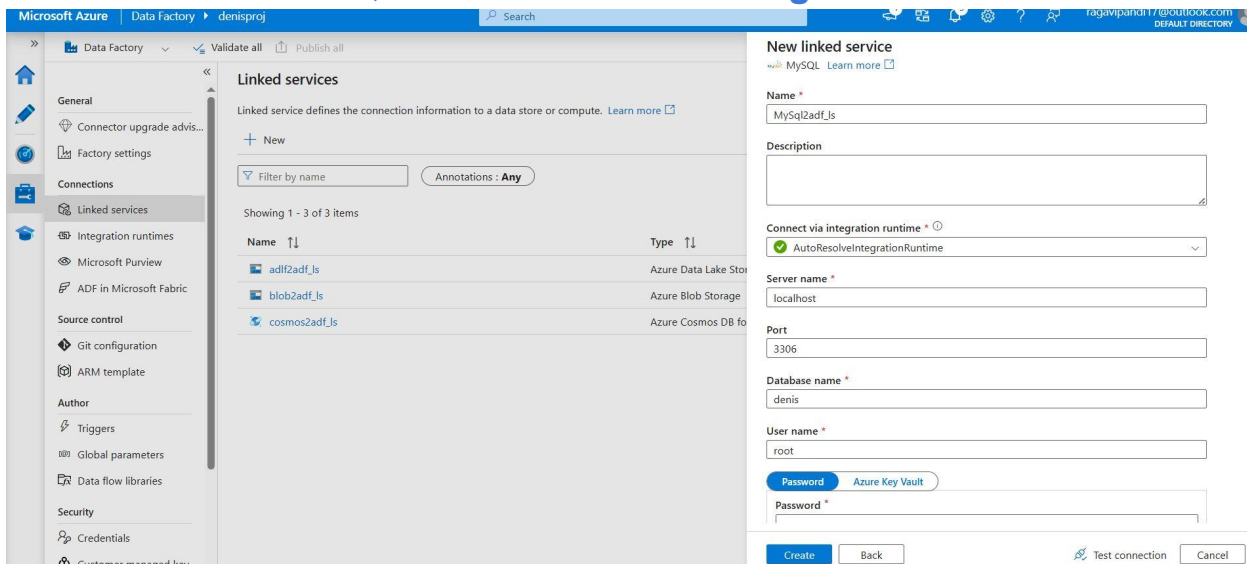


b.Before creating link service to **Onprem**, '**Self hosted integrated runtime**' should be installed in onprem device.

C. connect with onprem with Mysql Credentials



D. With other services, 'Give AutoResolveIntegrationRuntime'



Create Pipelines in ADF:

1.Raw2Bronze_PL:

Data move: Blob to ADLS

1.Create Pipeline in adf

2.Copy Activity

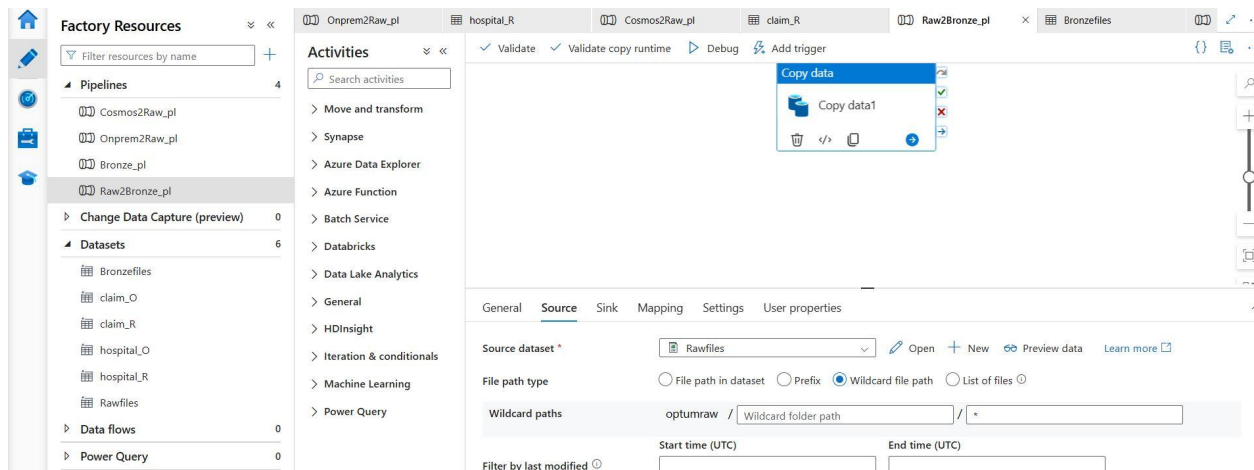
- Source : Blob, give '*' in wildcardpaths(to get all files from the container)

3.In datasets, select blob2adf_ls link service.

- Sink : Adls→Medallion→Bronze.

4.In datasets, select adls2adf_ls link service.

5.Trigger it and you will see the files in Bronze layer.



2. Cosmos2Bronze_PL:

Data move: Blob to ADLS

1. Create Pipeline in adf

2. Copy Activity

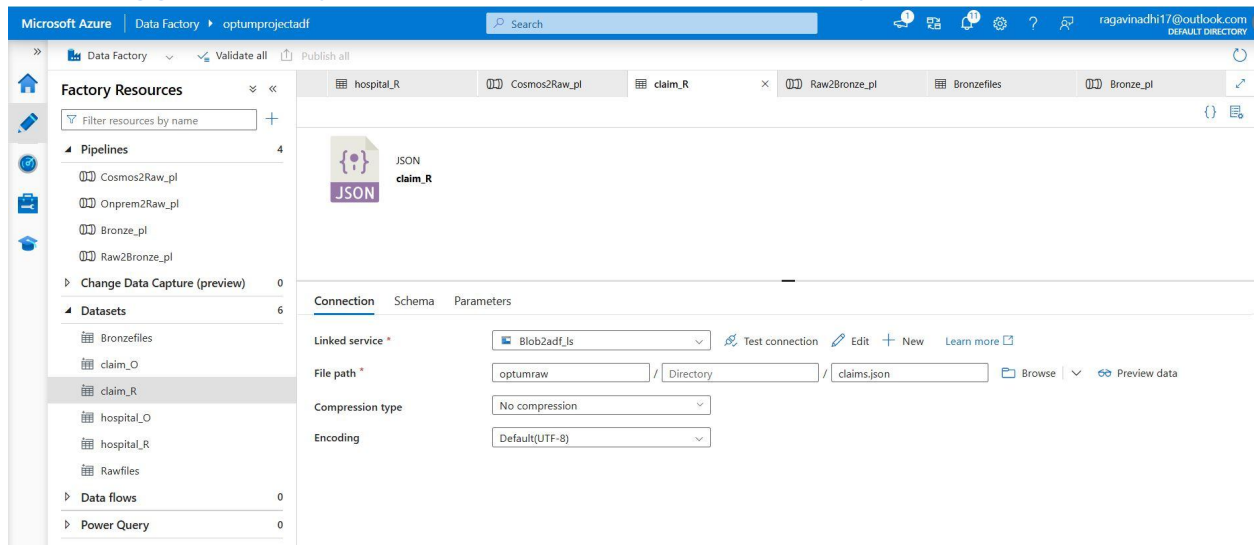
- Source : Cosmosdb.

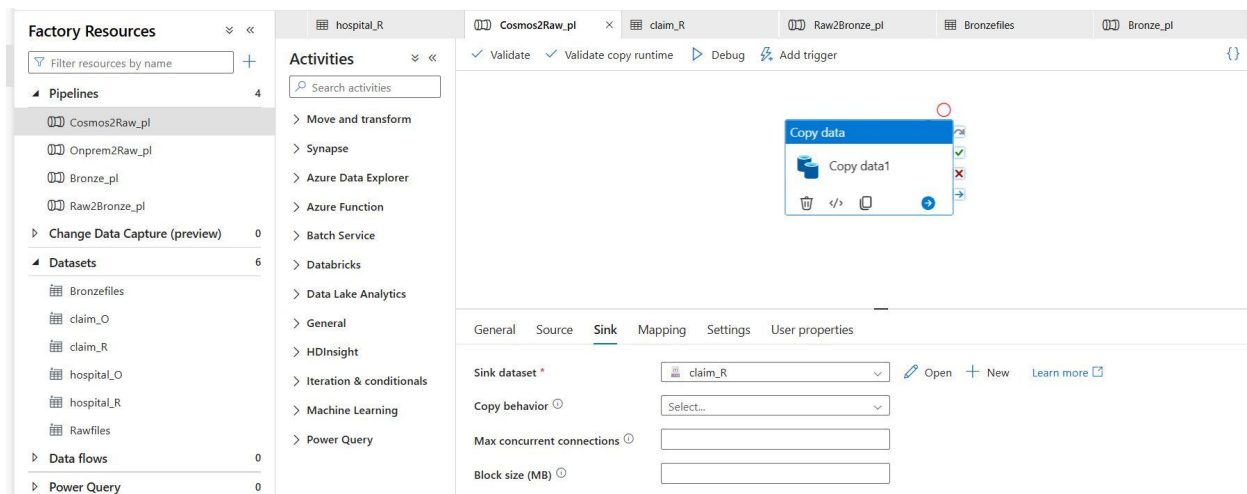
3. In datasets, select cosmosdb2adf_Is link service.

- Sink : Adls→Medallion→Bronze

4. In datasets, select adls2adf_Is link service.

5. Trigger it and you will see the files in Bronze layer.





3.Omnprem2Bronze_PL:

Data move: mysql to ADLS

1.Create Pipeline in adf

2.Copy Activity

- Source : MySQL-Onprem.

3.In datasets, select mysql2adf_Is link service.

- Sink : Adls→Medallion→Bronze

4.In datasets, select adls2adf_Is link service.

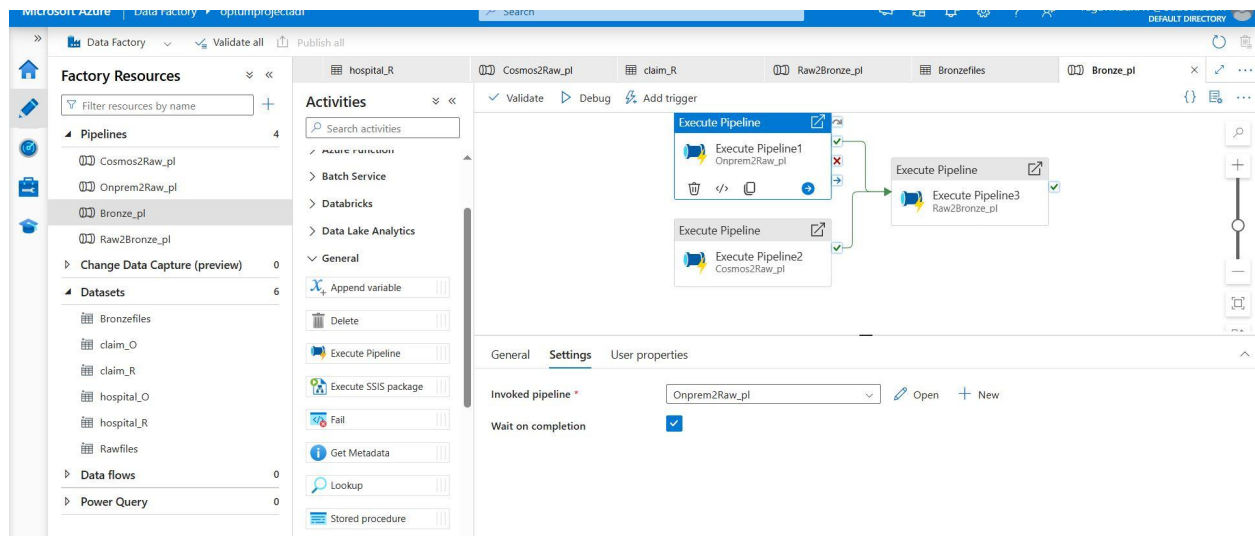
5. Trigger it and you will see the files in Bronze

layer.

The image displays two screenshots of the Azure Data Factory (ADF) interface. The top screenshot shows the 'Copy data' activity configuration in the 'Sink' tab, with the 'Sink dataset' set to 'hospital_R'. The bottom screenshot shows the 'Copy data' activity configuration in the 'Source' tab, with the 'Source dataset' set to 'hospital_O'. Both screenshots show the 'Factory Resources' pane on the left, listing pipelines and datasets. The 'Activities' pane on the right lists various activity types like 'Move and transform', 'Synapse', 'Azure Data Explorer', etc.

4. Create Bronze_PL:

Create 3 Execute Pipeline, Upload all 3 pipelines which needs to go Bronze layer.



5. Bronze2Silver data:

Data move: Adls(Bronze)to adls(Silver)

5.1. Transformation in Databricks:

Step 1:

Connect Adls to Databricks using either

- a. Access Token(Full Access)
 - b. Sas Token(Limited Access){read/write/delete etc}
 - c. Service Principal
2. Mount method(using above any of 3 methods)

Step 2.

1. Create Azure Key Vault

a. In Secret

- b. Name : AccessKeySecret
- c. Value: Password(From Blob storage)

2. And In Databricks, Give Scope Value

- a. Open Azure Secret Scope Documentations
- b. create secret scope in Databricks
- c. in Replace <Scope-name>, give scope name.
- d. Copy URL.
- e. duplicate the actual databricks url and paste the above copied url.

url.

Now,

f. Create Secret screen will open.

1. Give scope name.
2. DNS name : from key vault (Overview→Properties→copy vault uri) → paste it

3.Resource ID : from key vault (Properties → Copy Resource ID) → paste it

3.Copy scope name, paste it in Databricks<Scope>

All services > Key vaults >

Create a key vault

Basics **Access configuration** Networking Tags Review + create

Configure data plane access for this key vault

To access a key vault in data plane, all callers (users or applications) must have proper authentication and authorization. Authentication establishes the identity of the caller. Authorization determines which operations the caller can execute. [Learn more](#)

Permission model

Grant data plane access by using a [Azure RBAC](#) or [Key Vault access policy](#)

☐ Azure role-based access control (recommended) ⓘ

☒ Vault access policy ⓘ

Resource access

☐ Azure Virtual Machines for deployment ⓘ

☐ Azure Resource Manager for template deployment ⓘ

☐ Azure Disk Encryption for volume encryption ⓘ

Access policies

Access policies enable you to have fine-grained control over access to vault items. [Learn more](#)

[Previous](#) [Next](#) [Review + create](#)

Microsoft Azure Upgrade Search resources, services, and docs (G+/) Copilot ragavinadhi17@outlook... DEFAULT DIRECTORY (RAGAVINA...)

All services > optumsecurity | Keys ☆ ...

optumsecurity | Keys Key vault

Search Generate/Import Refresh Restore Backup Manage deleted keys

Overview Activity log Access control (IAM) Tags Diagnose and solve problems Access policies Resource visualizer Events Objects

Keys Secrets Certificates Settings Monitoring Automation Help

The operation is not allowed by RBAC. If role assignments were recently changed, please wait several minutes for role assignments to become effective.

Name	Status	Expiration date
You are unauthorized to view these contents.		

5.

2.Create Cluster in Databricks:

Microsoft Azure databricks Search data, notebooks, recents, and more... CTRL + P optumprodatabricks

Compute

All-purpose compute Job compute SQL warehouses Vector Search Pools Policies Apps Lakebase Postgres

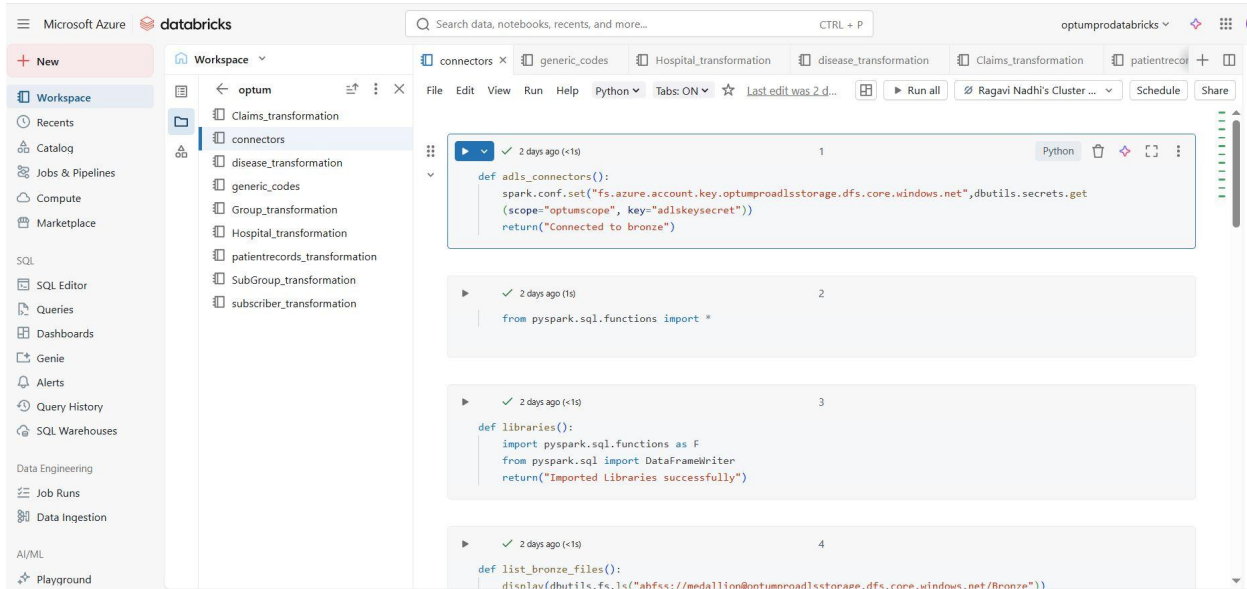
This account may not have enough CPU cores to start a cluster. Contact your administrator to increase the limits. Learn more about CPU quota.

Filter compute you have access to Created by Only pinned Create with Personal Compute Create compute

State	Name	Policy	Runtime	Active mem...	Active cores	Active DBU...	Source	Creator	Notebooks
●	Ragavi Nadhi's Cluster 2025-12-06 22:10...	-	16.4	16 GB	4 cores	2	UI	Ragavi Nadhi	-

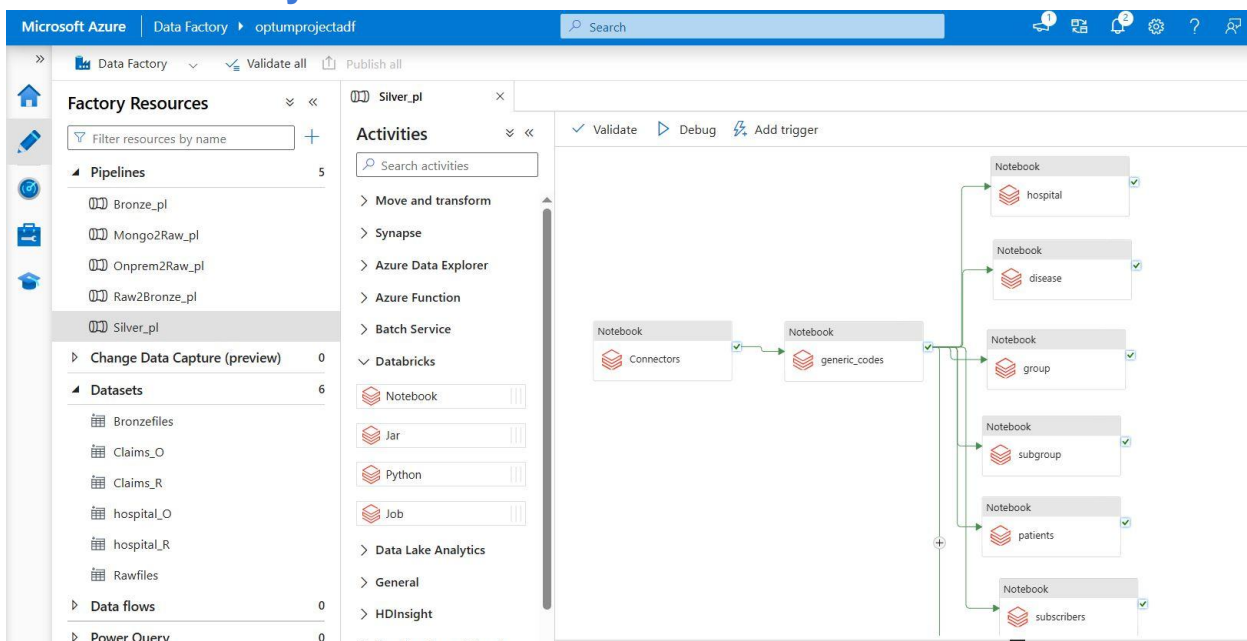
< Previous Next > 20 / page

5.3.Do all the Transformations in databricks



5.4. In adf, Create new pipeline, Silver_PL:

Create Notebooks as per the datasets, invoke it and connect them based on the continuity.



6. Silver2Gold Data:

1. Create one Pipeline name as Gold_PL
2. Create data flow name it as 'Optum_DF' and get Datas from silver layer
3. Connect them one by one with primaryID between the datas and finally get one single denormalised Data with required business calculations.
4. sink it to Medallion→Gold as a file and another sink in Azure SQL DB.

Microsoft Azure | Data Factory | optumprojectadf

Factory Resources

- Pipelines: 5
 - Bronze_pl
 - Mongo2Raw_pl
 - Onprem2Raw_pl
 - Raw2Bronze_pl
 - Silver_pl
- Change Data Capture (preview): 0
- Datasets: 13
 - Bronzefiles
 - Claims_O
 - Claims_R
 - Claims_S
 - Disease_S
 - Group_S
 - hospital_O
 - hospital_R
 - hospital_S

Silver_pl | optum_df

Validate | Data flow debug | Debug Settings

Sink Settings Errors Mapping Optimize Inspect Data preview

Output stream name: sink1

Description: Add sink dataset

Properties

General Related

Name: optum_df

Description:

Microsoft Azure | Data Factory | optumprojectadf

Factory Resources

- Pipelines: 5
 - Bronze_pl
 - Mongo2Raw_pl
 - Onprem2Raw_pl
 - Raw2Bronze_pl
 - Silver_pl
- Change Data Capture (preview): 0
- Datasets: 14
 - Bronzefiles
 - Claims_O
 - Claims_R
 - Claims_S
 - Disease_S
 - Group_S
 - hospital_O
 - hospital_R
 - hospital_S

Silver_pl | optum_df

Validate | Data flow debug | Debug Settings

Sink Settings Errors Mapping Optimize Inspect Data preview

Output stream name: sink1

Description: Add sink dataset

Incoming stream: allfilesjoins

Sink type: Dataset

Dataset: Select...

Options:

- Allow schema drift: ☒
- Validate schema: ☐

New dataset

In pipeline activities and data flows, reference a dataset to specify the location and structure of your data within a data store. Learn more

Select a data store:

Search

All Azure Database File Generic protocol

Azure Data Explorer (Kusto) Azure Data Lake Storage Gen2 Azure Database for MySQL

Azure Database for PostgreSQL Azure SQL Database Azure SQL Database Managed Instance

Continue Cancel

Microsoft Azure | Data Factory | optumprojectadf

Factory Resources

- Pipelines: 6
 - Bronze_pl
 - Mongo2Raw_pl
 - Onprem2Raw_pl
 - Raw2Bronze_pl
 - Silver_pl
 - Gold_pl
- Change Data Capture (preview): 0
- Datasets: 15
 - Bronzefiles
 - Claims_O
 - Claims_R
 - Claims_S
 - Disease_S
 - Group_S
 - hospital_O
 - hospital_R
 - hospital_S

Silver_pl | optum_df | Optum2asql | Gold_pl

Validate | Debug | Add trigger | Data flow debug

Activities

- Move and transform
 - Copy data
 - Data flow
- Synapse
- Azure Data Explorer
- Azure Function
- Batch Service
- Databricks
- Data Lake Analytics
- General
- HDInsight
- Iteration & conditionals
- Machine Learning
- Power Query

Data flow

optum_gold

Properties

General Related

Name: Gold_pl

Description:

Annotations

+ New

General Settings Parameters User properties

Name: optum_gold

Description:

Activity state: ☒ Activated ☐ Deactivated

7.Optum_PL:

- 1.Create one Pipeline ‘Optum_PL’
- 2.Create 3 execute_pipeline and invoke bronze_pl, silver_pl and gold_pl in here
- 3.Trigger ‘Optum_PL’, it will run all pipelines and we will see the final Output in AzureSqlDB and Gold layer in ADLS.

Home >

medallion

Container

Search

+

Add Directory

↑

Upload

↺

Refresh

🗑️

Delete

📄

Copy

📄

Paste

🔄

Rename

🔑

Acquire lease

🔑

Break lease

🔗

Edit columns

Overview

Diagnose and solve problems

Access Control (IAM)

Settings

medallion > Gold

Authentication method: Access key (Switch to Microsoft Entra user account)

Search blobs by prefix (case-sensitive)

Only show active objects

Showing all 1 items

<input type="checkbox"/>	Name	Last modified	Access tier	Blob type	Size	Lease state
<input type="checkbox"/>	[-]					...
<input type="checkbox"/>	optum_G.csv	12/12/2025, 11:45:00 PM	Hot (Inferred)	Block blob	488.47 KiB	Available

Add or remove favorites by pressing Ctr-L+Shift+F